THE BATTLE OF NEIGHBORHOODS

BY REGINALD DUBE

Motivation

- Suppose a person wants to move from New York to Toronto for a job. This person does not know anything about Toronto and he/she would like to move to a place similar to where he/she currently lives.
- Is it possible to create a system that can help our user showing ti him/her the similarities between the two countries.

Objectives

- Develop a system able to show similarities in terms of neighborhoods in order to help a user decide whether to move near the center of Toronto or not.

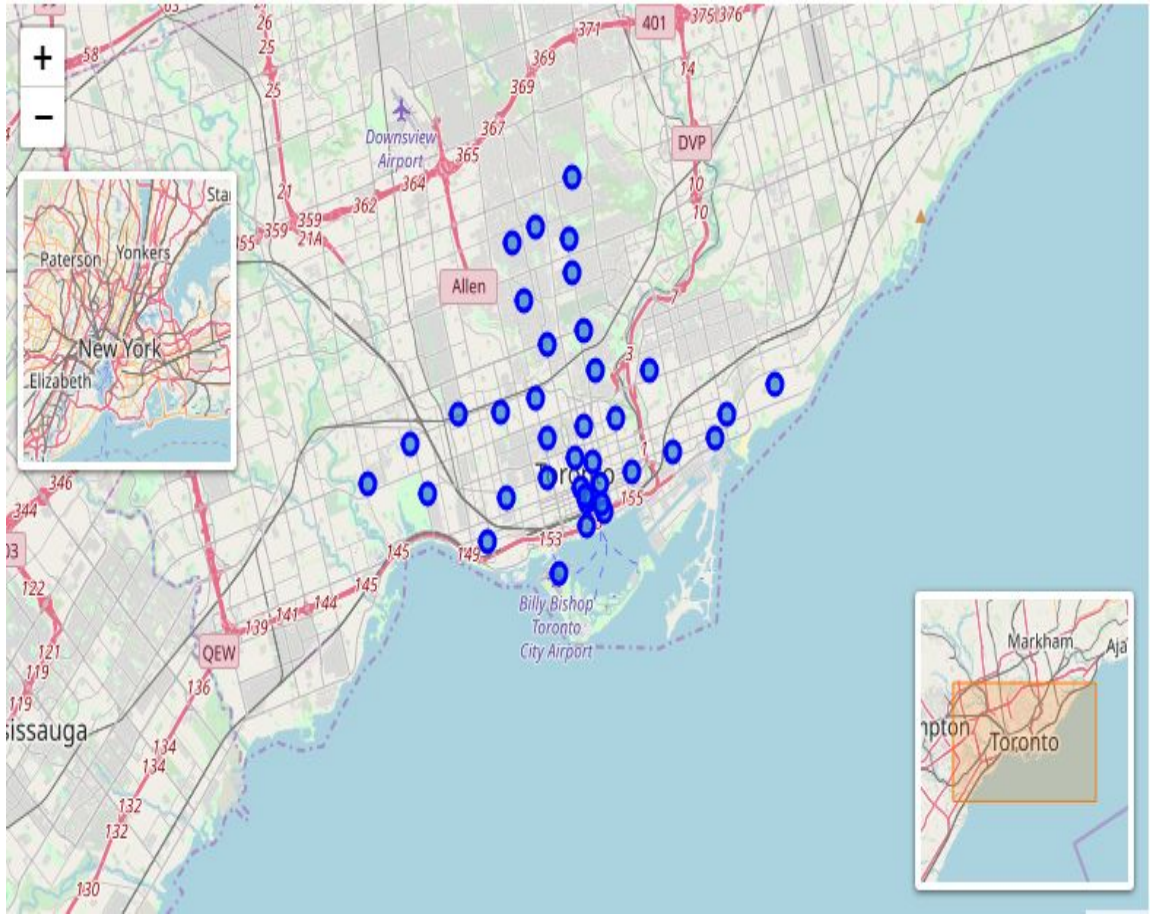
PROPOSAL

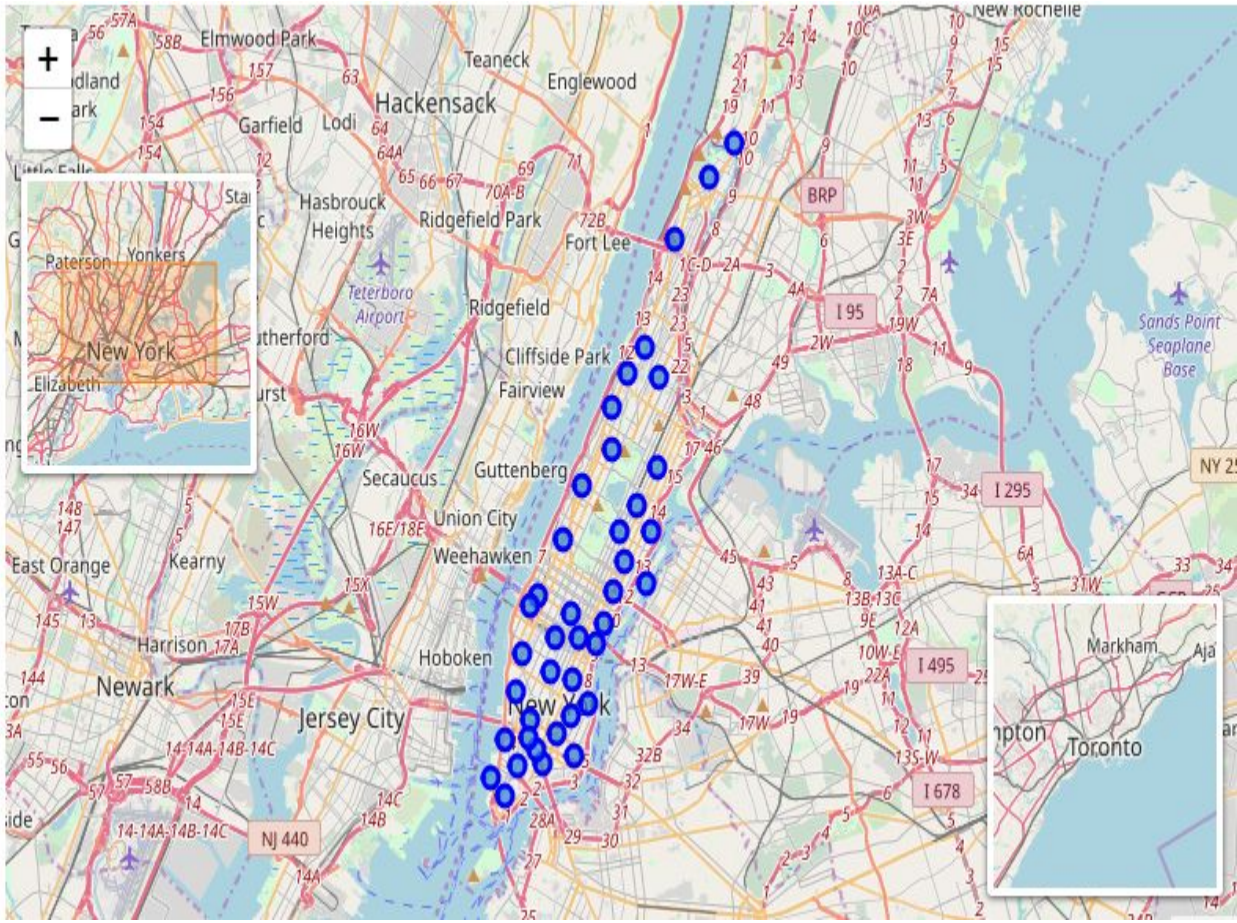
Approach

- Neighborhoods are downloaded
- Venues are requested using Foursquare API
- The categories of venues are encoded using One Hot
- K-means algorithm is used for finding similarities
- The elbow method is used for select K

RESULTS

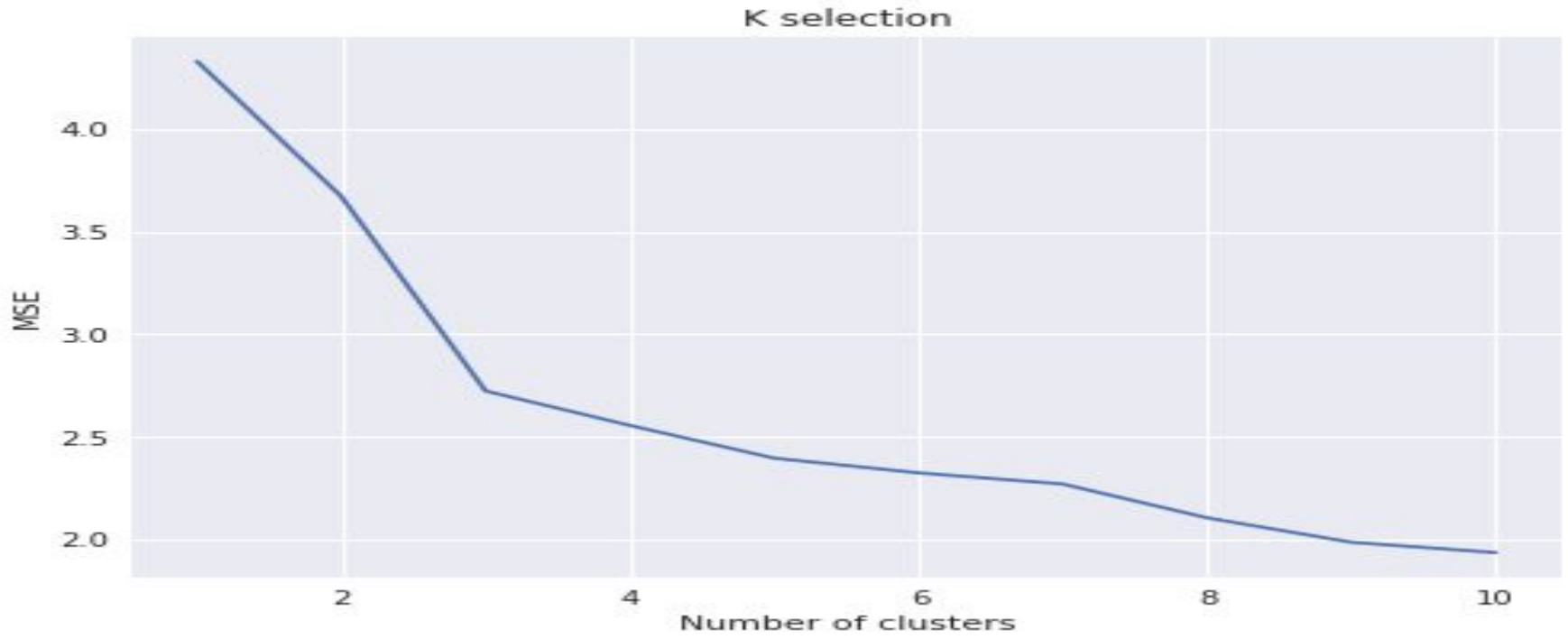
GEOGRAPHICAL LOCATION



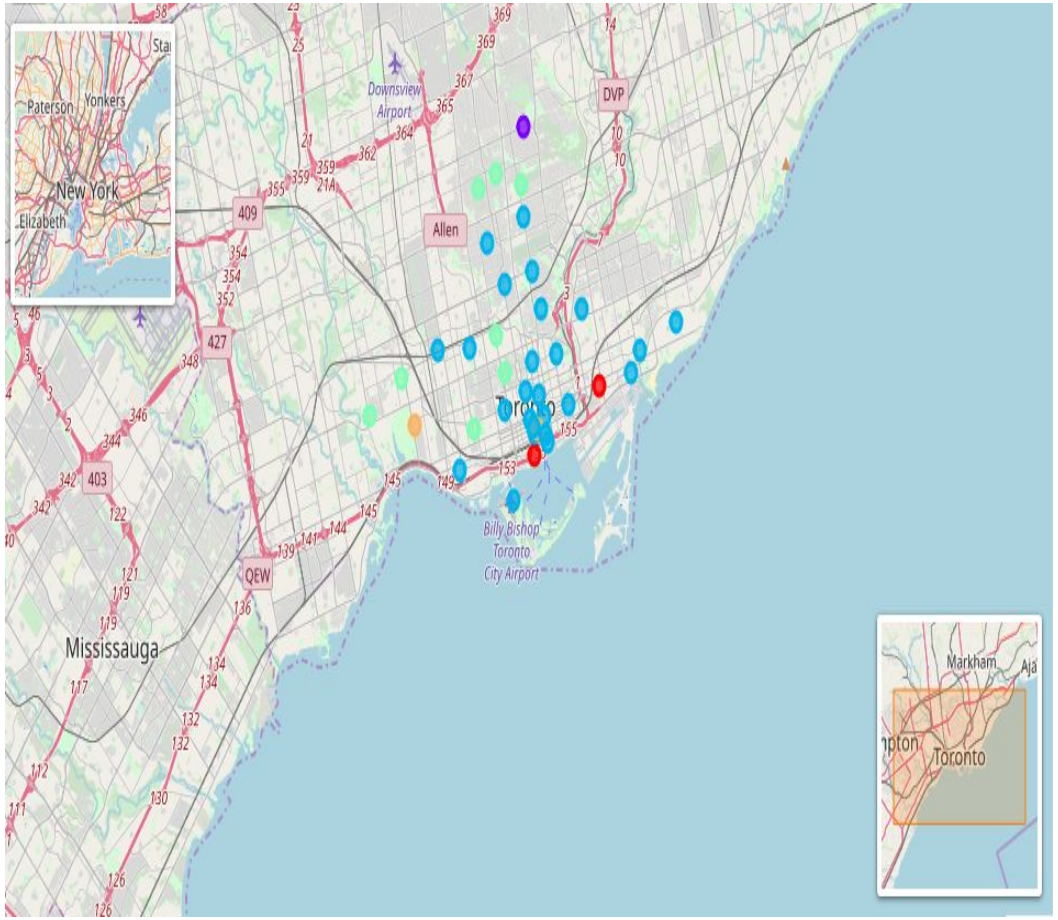


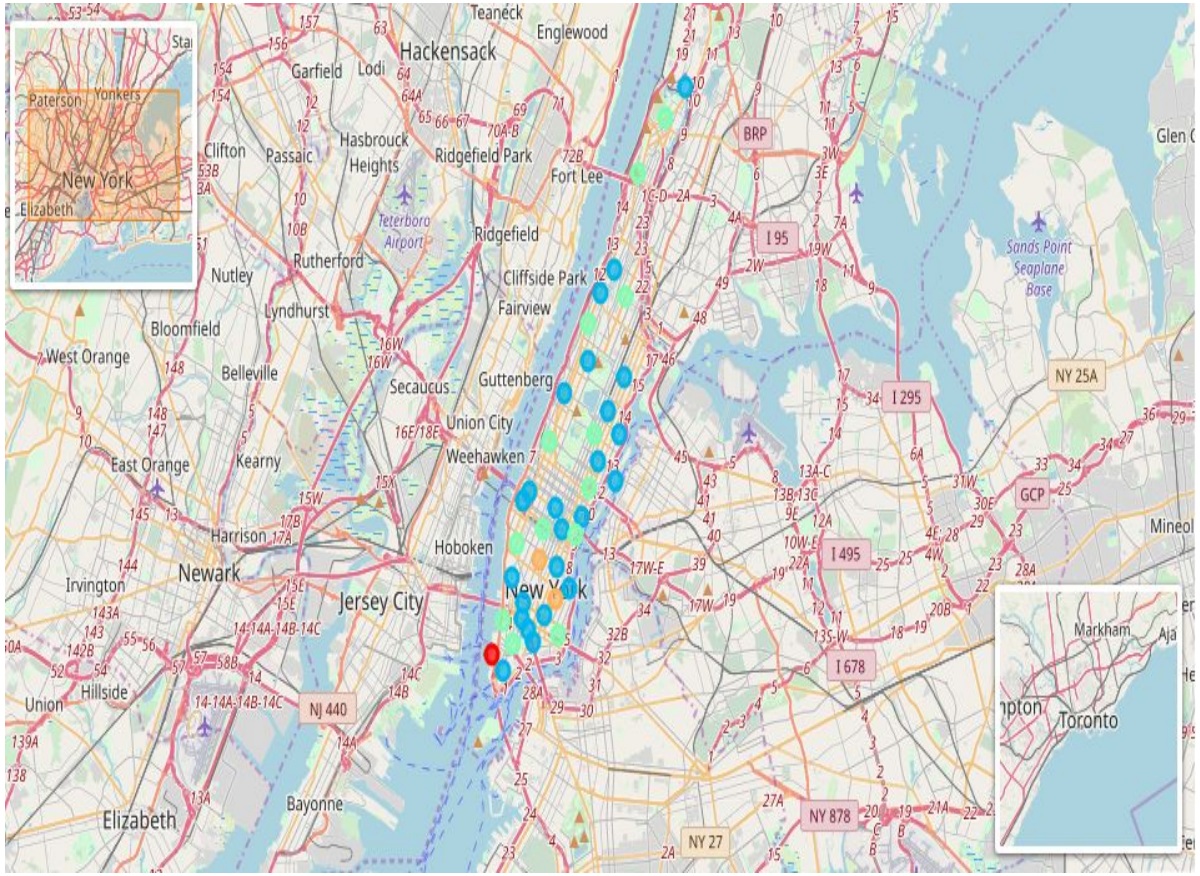
Selection of K

- The best number of cluster is 5. That is , where the elbow is located. After that, the mean squared error decrease without big changes.

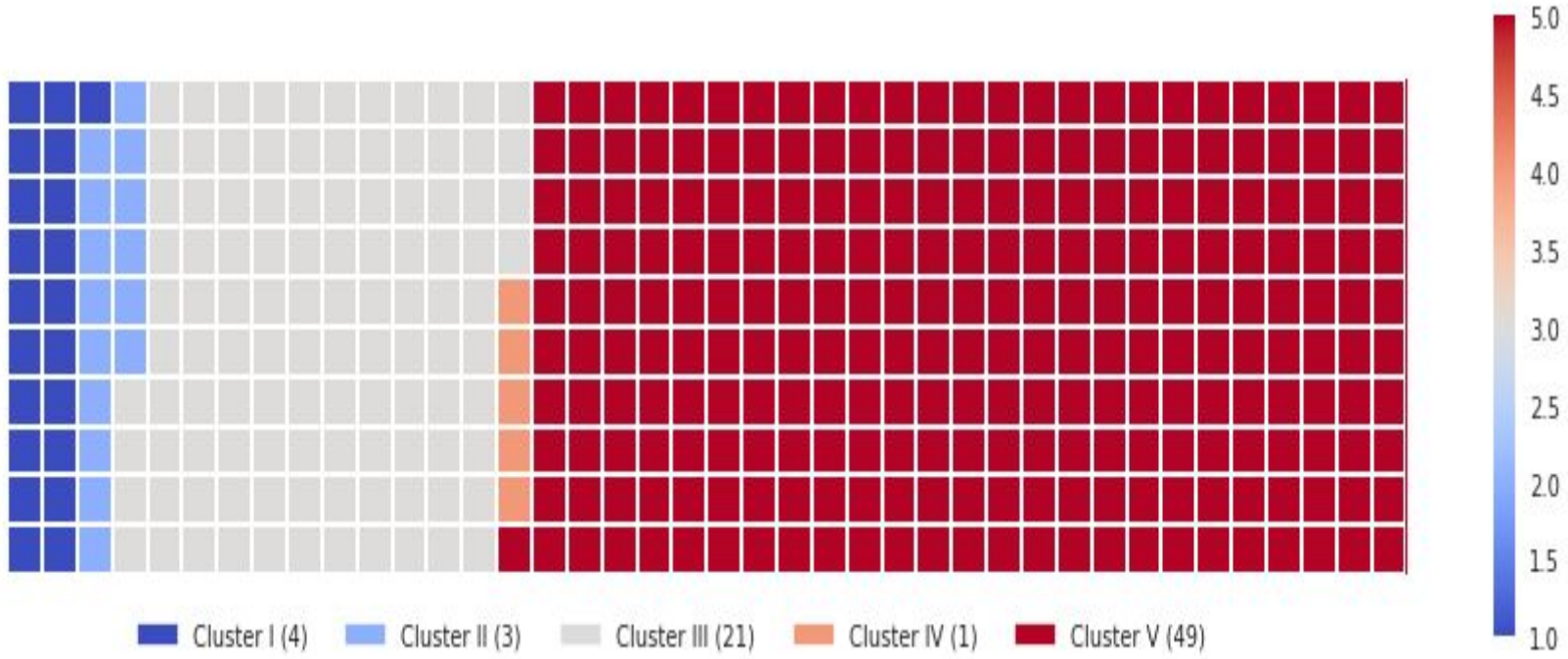


Geographical Location(Clustered)

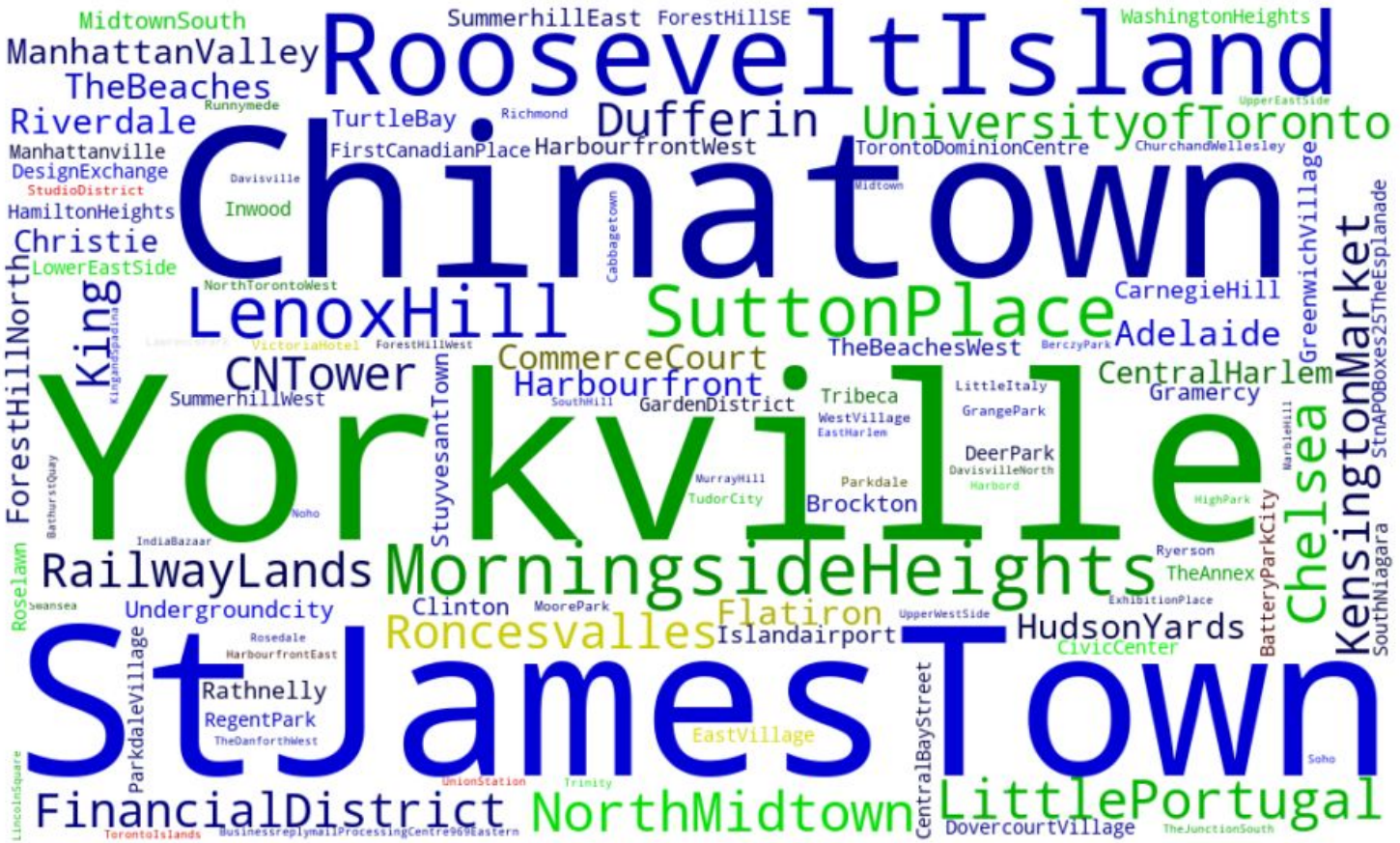




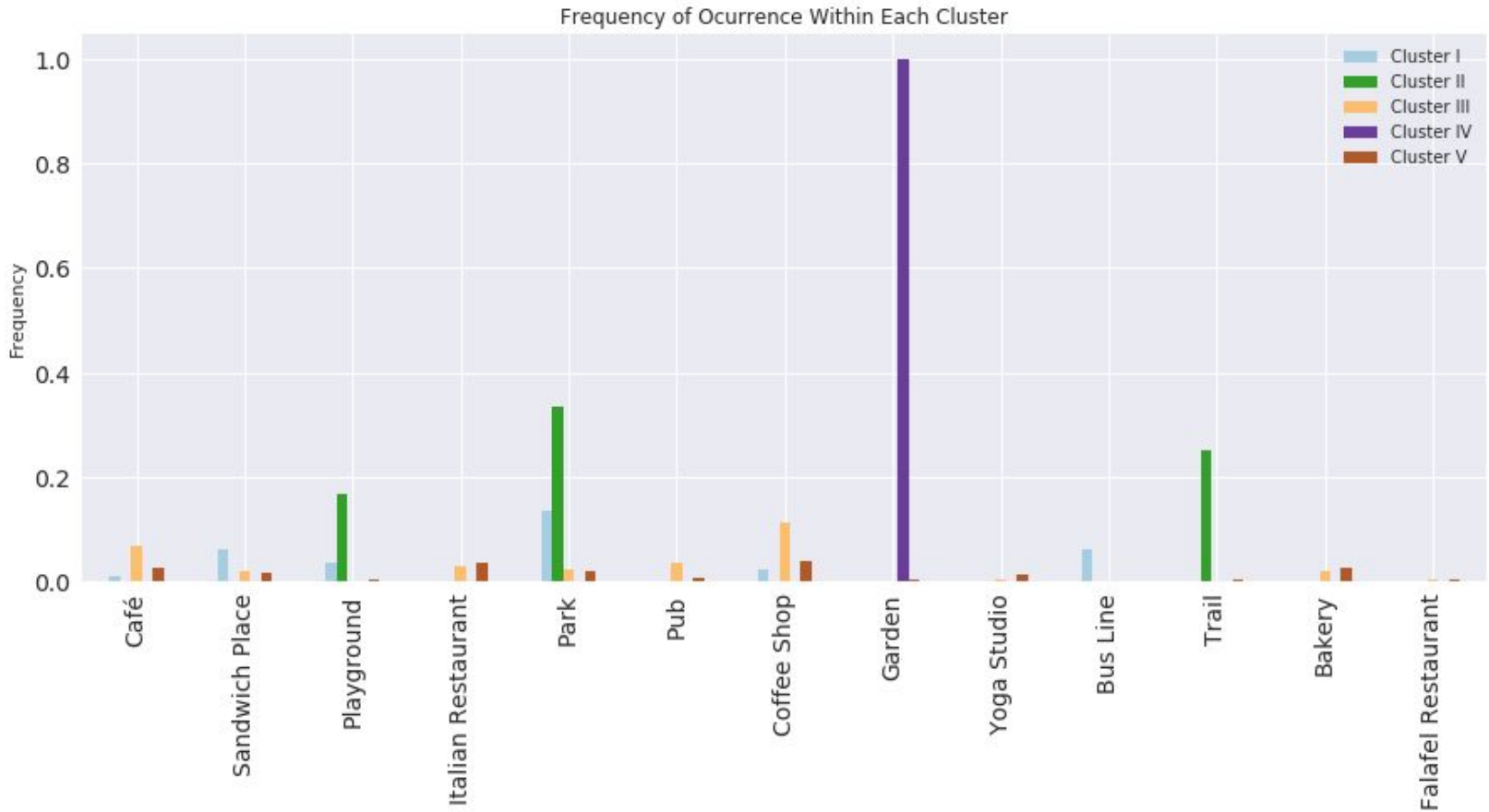
Proportion of Data Segmented



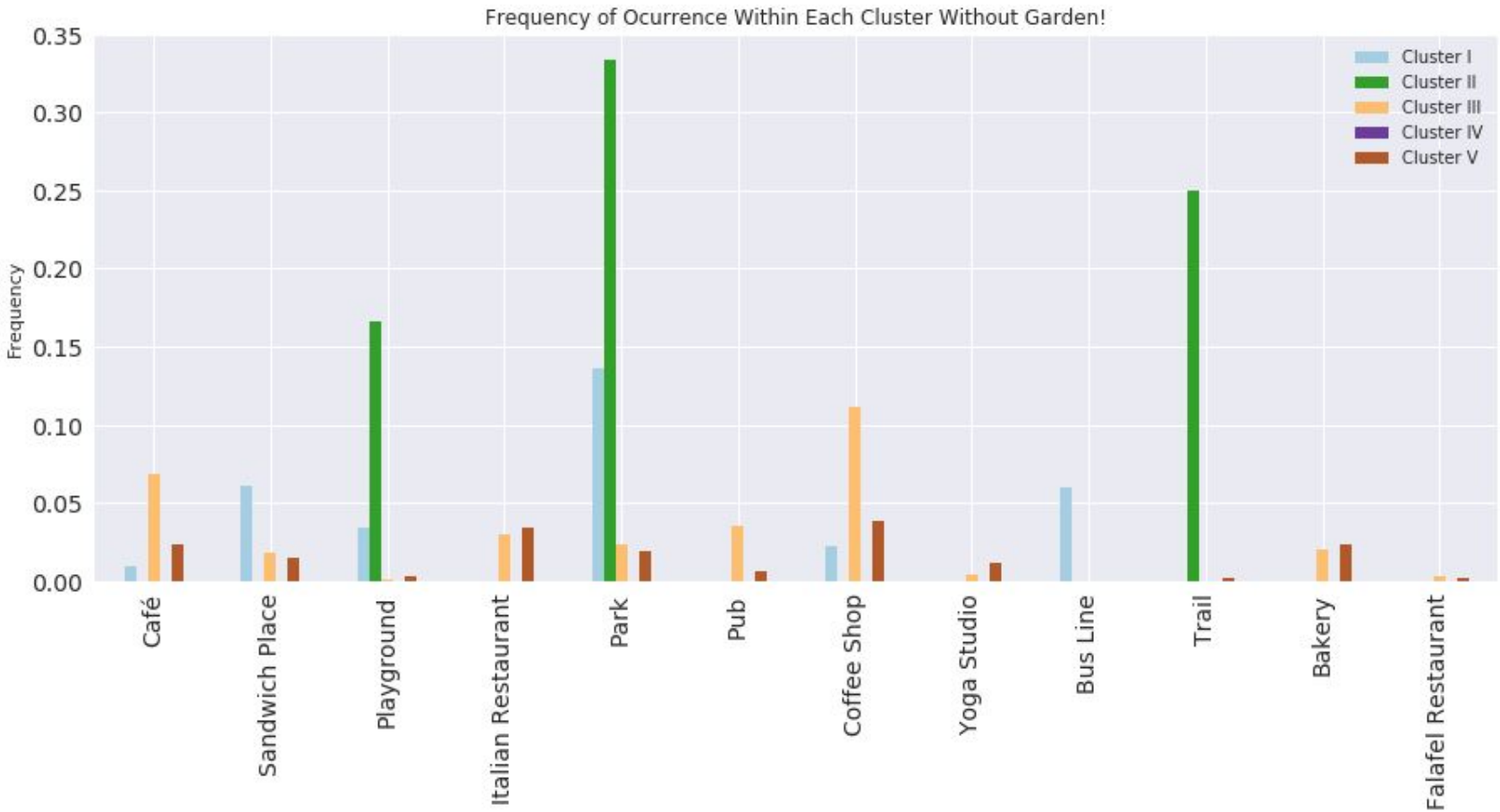
Neighborhoods Segmented By Colors



Bar Chart (Frequent Venues)



Bar Chart (Without Garden)



Conclusion

- i) Neighborhoods that have parks around , bus lines and coffee shops
- ii) Neighborhoods that have pubs and Italian restaurants etc